DM-Group Meeting

Liangzhe Chen, Jan. 22 2015
Papers to be present

- SNOC: Streaming Network Node Classification
  - ICDM 2014, by Ting Guo et. al.

- Scalable SVM-based Classification in Dynamic Graphs
  - ICDM 2014, by Yibo Yao et. al.

- Multi-Graph-View Learning for Graph Classification
  - ICDM 2014, by Jia Wu et. al.
SNOC: Streaming Network Node Classification

ICDM 2014
Ting Guo, Xingquan Zhu, Jian Pei, and Chengqi Zhang
Motivation

- Networks are changing
  - New nodes are added.
  - New edges are created.
  - Node contents are changed.

- How can nodes be classified in such dynamic/streaming network?
Streaming network node classification aims to classify unlabeled nodes in the network, at any time $t$, with maximal accuracy.
Proposed Method

- The theme is to let:
  - Nodes sharing the same class and having a high structure similarity be close to each other
  - Nodes belonging to different classes and having a weak structure relationship be far away from each other

\[ y^{u*} = \arg \min_{y^u \subseteq y} E(y^u) \]
Proposed Method

- Find a good set of features to capture the changes in the network.

\[ y^u_* = \arg \min_{\forall y^u \in Y} E(y^u, S) \]
How to find $S$? It should take into account:

- The label-based node similarity in the Label Space
- The structure-based node similarity in the Structure Space

Laplacian based quality criterion

$$E(Y^u, S) = \frac{1}{2} \sum_{i \in X^u} \sum_{j \in X} h(i, j, y_i)(D_S x_i - D_S x_j)^2$$

subject to

$$\min \left( \frac{1}{2} \sum_{i, j \in X} h(i, j)(D_S x_i - D_S x_j)^2 \right), S \subseteq F, |S| = m$$
Experiments: Datasets

<table>
<thead>
<tr>
<th>Data sets</th>
<th># Nodes</th>
<th># Edges</th>
<th># Features</th>
<th># Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cora</td>
<td>2,708</td>
<td>5,429</td>
<td>1,433</td>
<td>7</td>
</tr>
<tr>
<td>CiteSeer</td>
<td>3,312</td>
<td>4,732</td>
<td>3,703</td>
<td>6</td>
</tr>
<tr>
<td>PubMed Diabetes</td>
<td>19,717</td>
<td>44,338</td>
<td>500</td>
<td>3</td>
</tr>
<tr>
<td>DBLP</td>
<td>2,084,055</td>
<td>2,244,018</td>
<td>3,000</td>
<td>6</td>
</tr>
</tbody>
</table>
Experiments: Results

(a) Cora  (b) CiteSeer  (c) PubMed Diabetes

Accuracy %

# of selected features (m)
Experiments: Results

(a) Maximal Path Length $l$

(b) Weight Parameter $\xi$

(c) Percentage of Labeled Nodes
Experiments: Results

(a) DBLP

(b) PubMed Diabetes
Experiment: Results

(a) DBLP

(b) PubMed Diabetes

(c) extended DBLP
Scalable SVM-based Classification in Dynamic Graphs

ICDM 2014

Yibo Yao and Lawrence Holder
Fig. 1. An example of a dynamic graph. Updates are received in the form of batches. For example, when $B_2$ comes in, case 2 happens since two new nodes $F, H$ are connected to two old nodes $B, E$; case 4 also happens because a new edge connecting two new nodes $H, G$ is inserted.
Fig. 2. An instance of a citation network. $P_i$ represents a paper of interest while $A_i$ represents an author associated to a paper. An edge label represents the relationship between the two nodes it connects.
Problem Definition

- Given a dynamic graph with central and side nodes, and each central node $v_i$ has an associated class label $y_i$ (+1/-1)
- Learn a classifier using available information up until the current time $t$, and predict the class labels of new central nodes from the next batch.
Framework of Solution

- When a new batch $B_t$ comes
  - Extract subgraphs for the central nodes in $B_t$ (entropy-based method)
  - If a sliding window is specified, delete the old information which is outside the current window
  - Combine the support vectors of the classification model learned from $B_{t-1}$ with the subgraphs from $B_t$ as a new training set, and learn a new model to predict the class labels of the central nodes from $B_{t+1}$
Algorithm 2 Incremental SVM (IncSVM)

**Input:**
- $\mathcal{G}$: A graph
- $SV_{t-1}$: A set of support vectors
- $B_t$: The current batch
- $\theta$: The threshold for selecting neighbor nodes

**Output:**
- $M_t$: A SVM classification model for prediction

1. $Sub_{B_t} = \emptyset$
2. **for** each central node $v_c$ in $B_t$ **do**
3. \quad $Sub_{B_t} = Sub_{B_t} \cup \{\text{SubExtract}(\mathcal{G}, v_c, \theta)\}$
4. **end for**
5. construct a training set $TR_t = SV_{t-1} \cup Sub_{B_t}$
6. learn a classifier $M_t$ on $TR_t$ using W-L kernel
7. **return** $M_t$ (including its support vectors $SV_t$)
Experiment: Datasets

- IMDB Network: The Internet Movie Database. The task is to predict whether a new movie will be successful.
- DBLP Network: The task is to predict whether a paper belongs to DBDM or CVPR
Experiment: Results

Fig. 8. Average accuracy across all batches and accumulated learning time on IMDB w.r.t. different values of $\theta$. 
Experiment: Results

Fig. 9. Average accuracy across all batches and accumulated learning time on DBLP w.r.t. different values of $\theta$. 
Multi-Graph-View Learning for Graph Classification

ICDM 2014

Jia Wu, Zhibin Hong, Shirui Pan, Xingquan Zhu, and Chengqi Zhang
There are different channels/views to describe objects, resulting in a new representation with multiple graphs generated from different feature views being used to describe one object.
Motivation
Definition 4: (Subgraph) Let $G = (\mathcal{V}, E, \mathcal{L}, l)$ and $g_i = (\mathcal{V}', E', \mathcal{L}', l')$ each denote a connected graph. $g_i$ is a subgraph of $G$, i.e., $g_i \subseteq G$, iff there exists an injective function $\varphi : \mathcal{V}' \rightarrow \mathcal{V}$ s.t. (1) $\forall v \in \mathcal{V}', l'(v) = l(\varphi(v))$; (2) $\forall (u, v) \in E'$, $(\varphi(u), \varphi(v)) \in E$ and $l'(u, v) = l(\varphi(u), \varphi(v))$. If $g_i$ is a subgraph of $G$, then $G$ is a supergraph of $g_i$. In this paper, subgraphs and subgraph features are equivalent terms.

Definition 5: (Graph Feature Representation) Let $S_k = \{g_1, \cdots, g_{s_k}\}$ denote a set of subgraph features discovered from multi-graph-view graphs. For each graph $G_k^i$ in the $k^{th}$ view, we use a subgraph feature vector $x_i^k = [(x_i^{g_1})^k, \cdots, (x_i^{g_{s_k}})^k]^{\top}$ to represent $G_k^i$ in the feature space, where $(x_i^{g_e})^k = 1, 1 \leq e \leq s_k$, iff $g_e$ is a subgraph of $G_k^i$ (i.e., $g_e \subseteq G_k^i$) and $(x_i^{g_e})^k = 0$ otherwise.
Goal:

To find the optimal subgraph features from the training graph set $G$ to train classification models, and predict previously unseen multi-graph-view graphs with a maximum accuracy.
Subgraph Evaluation Criteria:
- Find informative-irredundant subgraph features across all graph-views

Cross Graph-View Subgraph Selection:
- Automatically assign weight values to different graph-views, and further optimize the weight values to ensure that high quality subgraphs are selected from important graph-views.

Multi-Graph-View Graph Representation:
- Concatenate subgraph features across all graph-views to form the final graph representation
Experiment: Datasets

- DBLP
- Images from Corel dataset
Figure 3. Comparisons on *DBLP dataset* on each single graph-view: (A) Reference Relationship view; (B) Abstract view.
Figure 4. Comparisons on DBLP dataset with multiple graph-views by using different view combination strategies.
Figure 6. Comparisons on *Image dataset* with multiple views by using different view combination strategies.